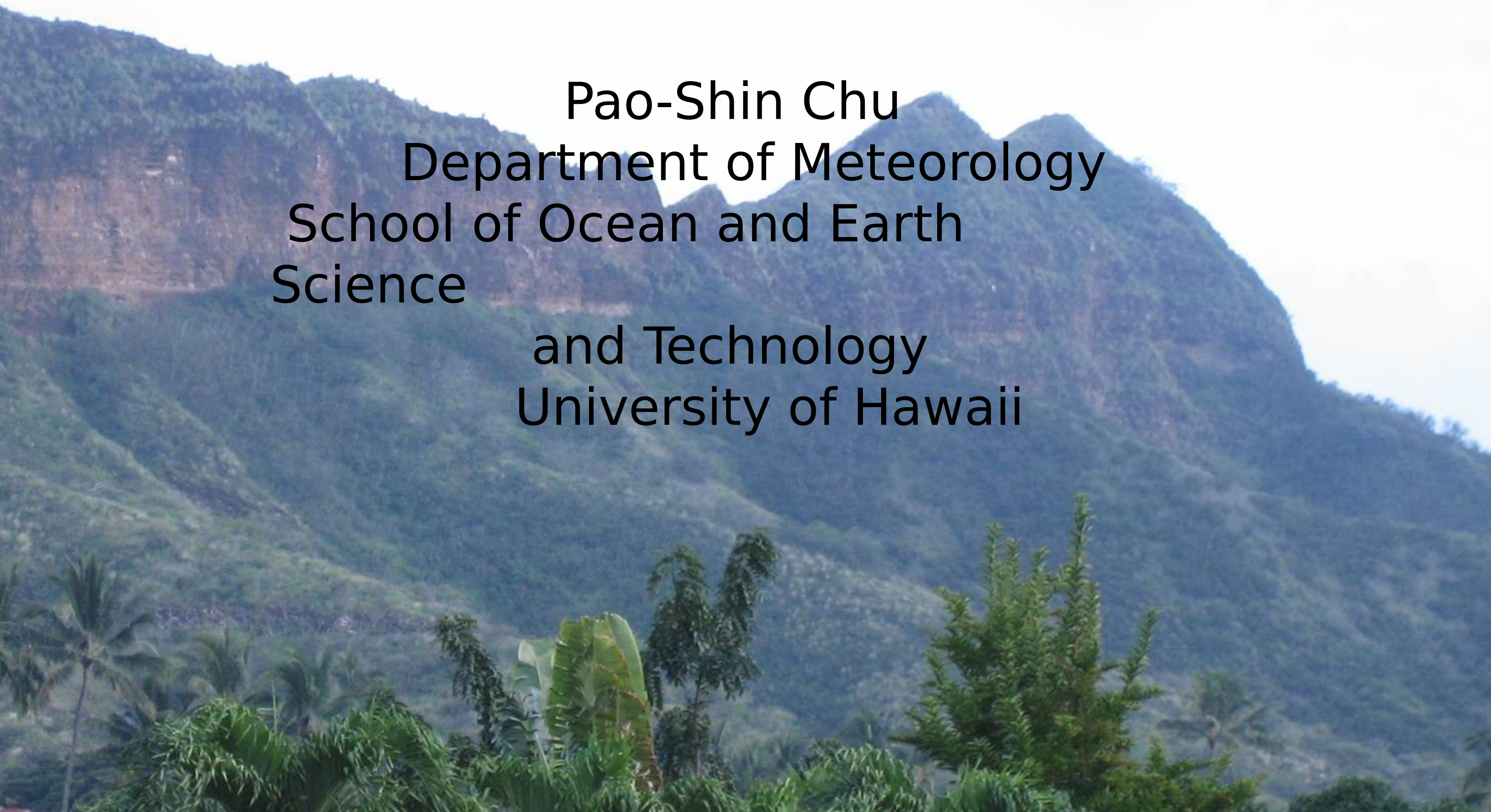
Forecasting Seasonal and Regional Typhoon Activity: A Track Based Approach

Pao-Shin Chu
Department of Meteorology
School of Ocean and Earth Science
and Technology
University of Hawaii



- X. Zhao (UH)
- C.-H. Ho, H.-S. Kim, and J.-H. Kim
  (Seoul National University)
- M.-M. Lu (Central Weather Bureau, Taiwan)

Part I: Probabilistic Forecasting (Bayesian)

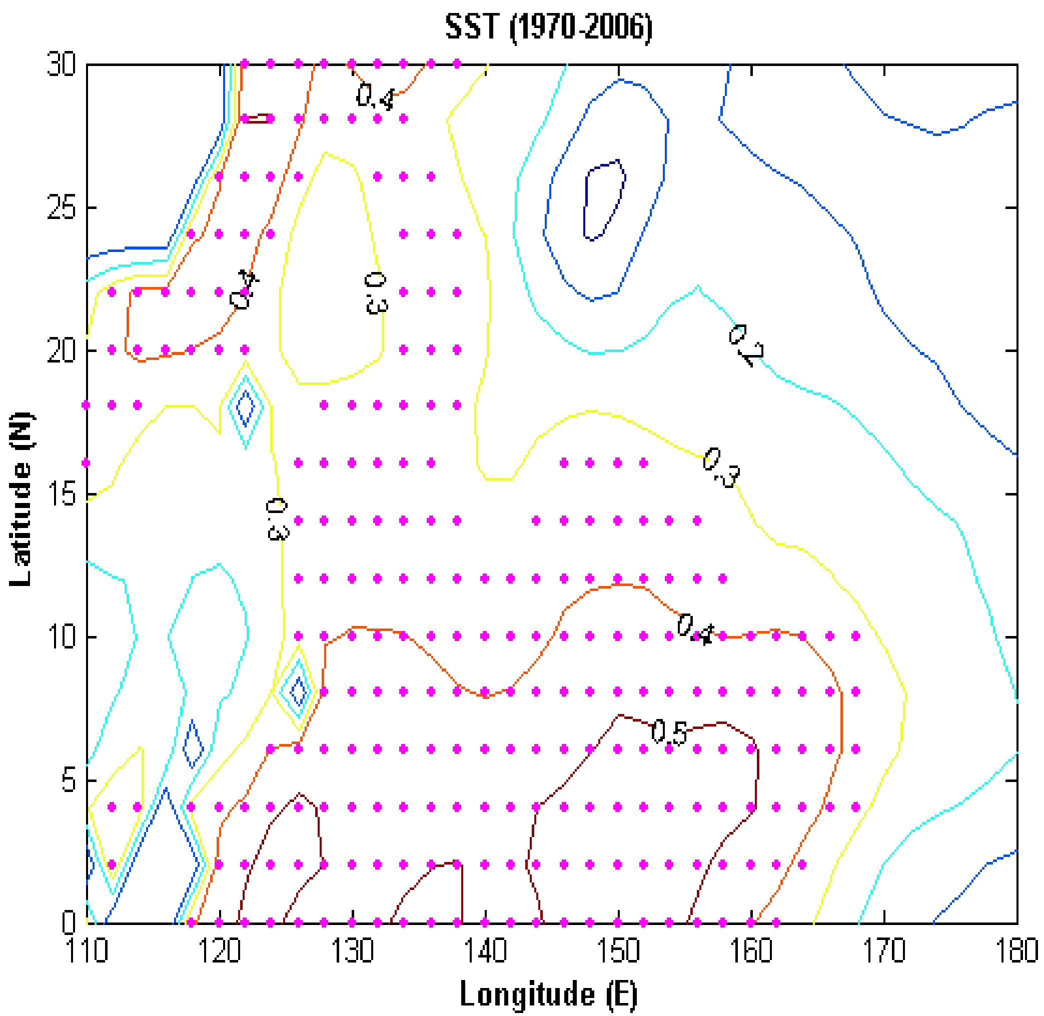
Part II: Track-type Based Probabilistic Forecasting

- Basin-wide (Chan et al. 1998 & 2001, Wea. Forecasting) to **regional** forecasts (Chu et al., 2007, TAO) for adequate planning of regional emergency management and hazard mitigation

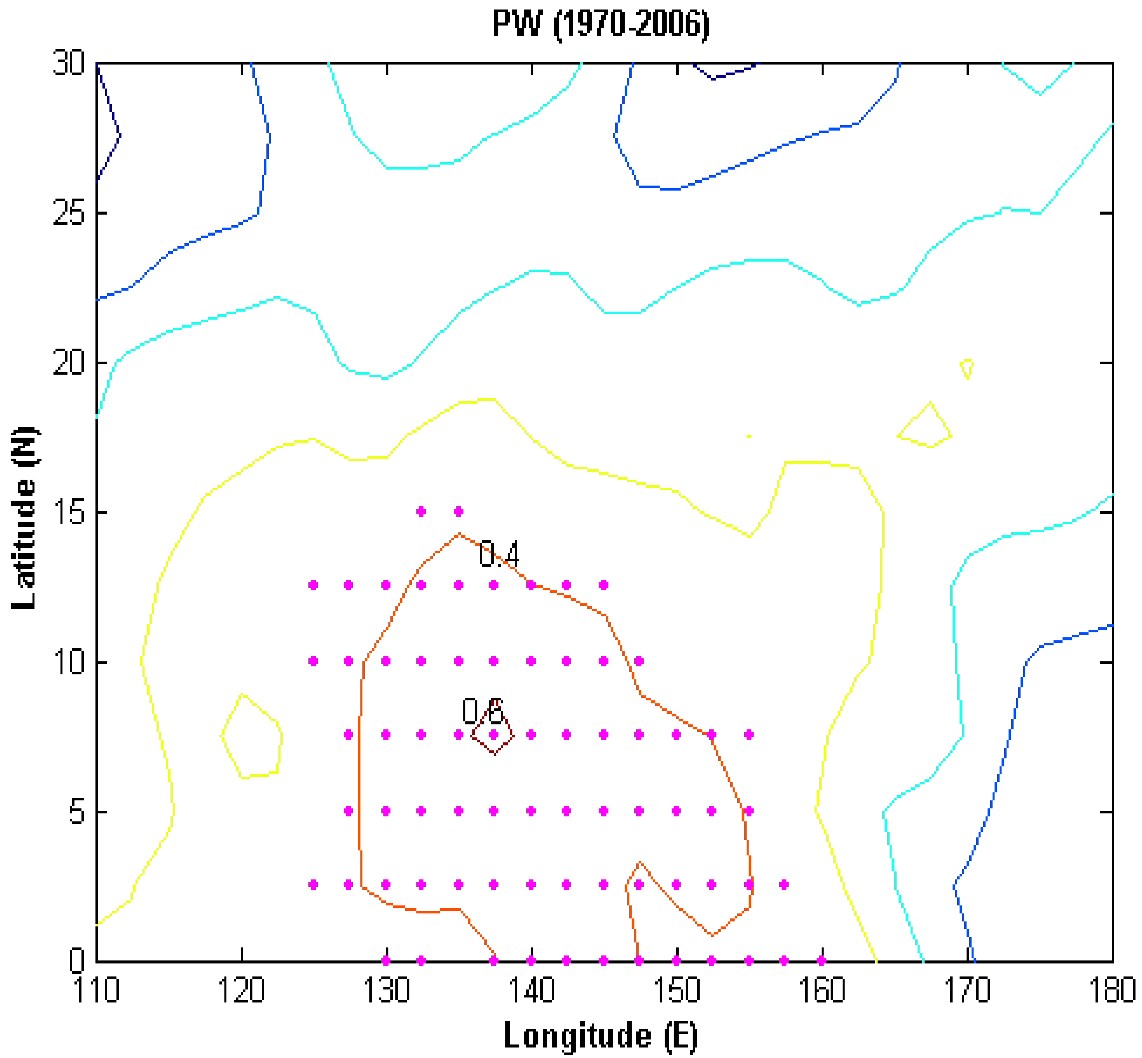
- Deterministic (Chan et al., 1998, 2001; Chu et al., 2007) to **probabilistic** (Chu and Zhao, 2007, J. Climate) to facilitate decision-making

- Tropical cyclone data in the vicinity of Taiwan (21-26°N, 119-125°E) from 1970 to 2006 compiled by the CWB in Taiwan

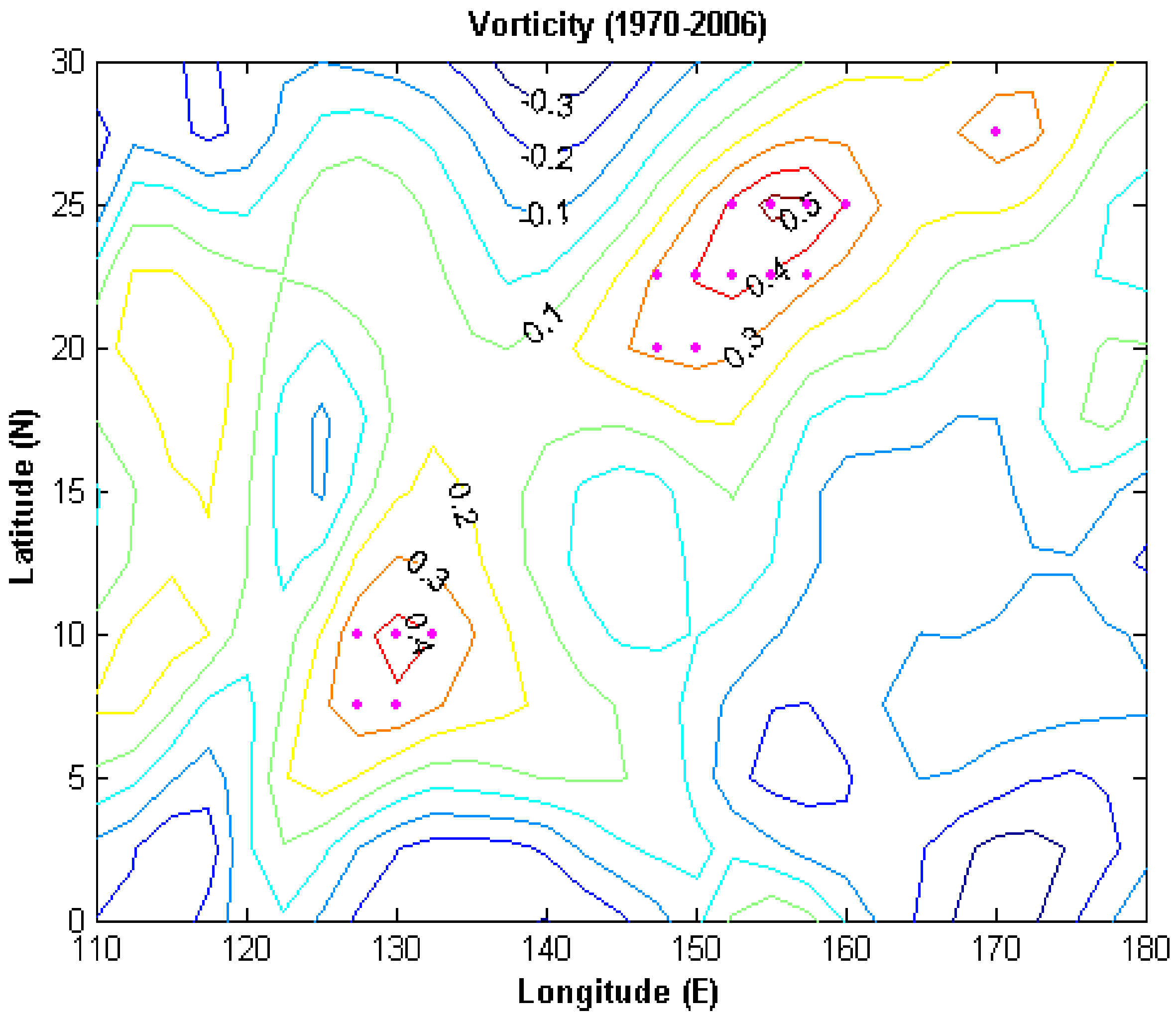
- Monthly mean SST, wind data at 850- and 200-hPa levels, relative vorticity at the 850 hPa level, and total precipitable water over the tropical western North Pacific (NCEP/NCAR Reanalysis products)



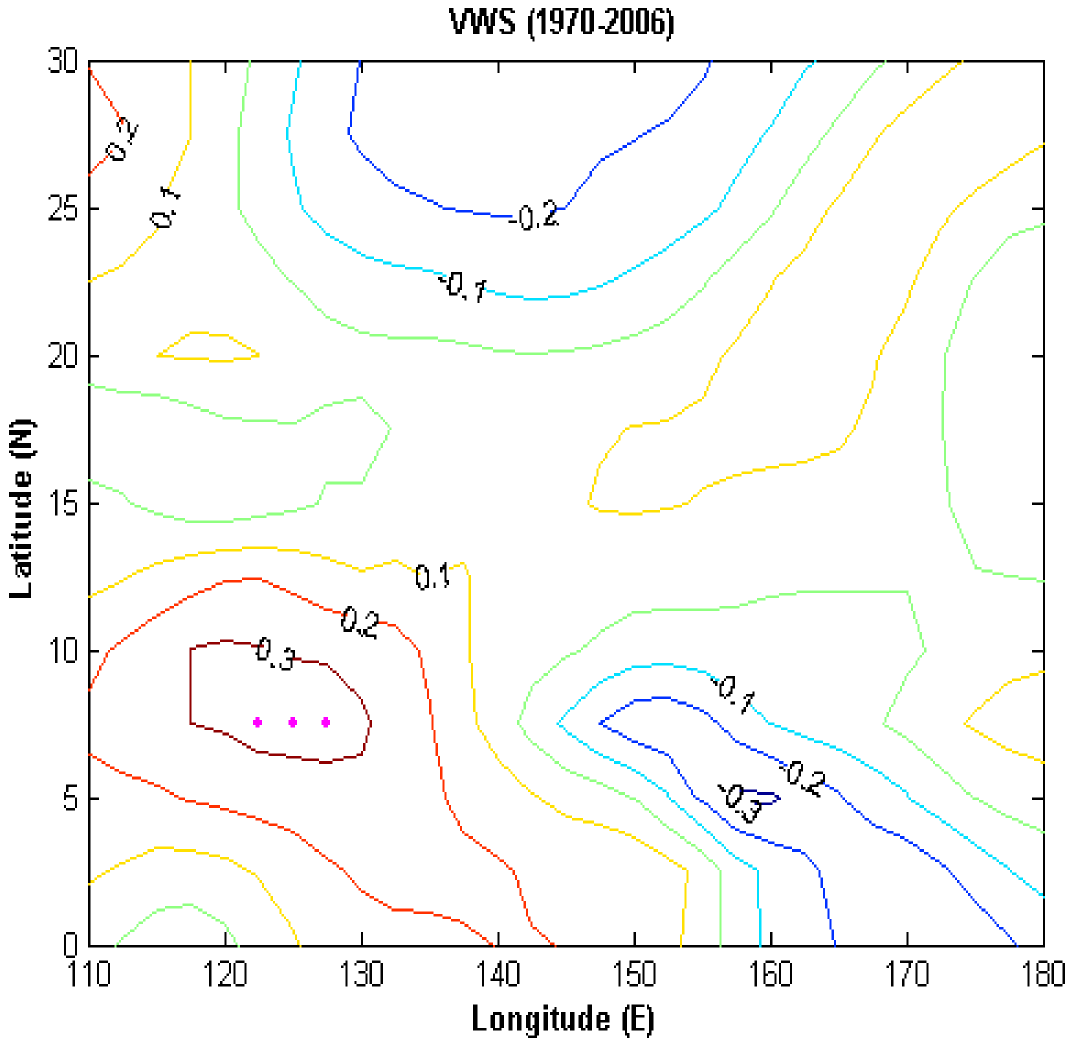
SST (1970-2006)
Latitude (N)
Longitude (E)
30
25
20
15
10
5
0
110
120
130
140
150
160
170
180
0.4
0.3
0.2
0.5



PW (1970-2006)
Latitude (N)
Longitude (E)
0.4
0.6
30
25
20
15
10
5
0
110
120
130
140
150
160
170
180



Vorticity (1970-2006)
Latitude (N)
Longitude (E)
30
25
20
15
10
5
0
110
120
130
140
150
160
170
180
-0.3
-0.2
-0.1
0.1
0.2
0.3
0.4
0.5



VWS (1970-2006)
Latitude (N)
Longitude (E)
0
5
10
15
20
25
30
110
120
130
140
150
160
170
180
0.2
0.1
-0.2
-0.1
0.1
0.2
0.3
-0.1
-0.2
-0.3

Part I: Probabilistic Forecasting

Poisson Distribution

Given the Poisson intensity parameter λ (i.e., the mean seasonal TC rates), the probability mass function (PMF) of h TCs occurring in T years is (Epstein, 1985)

$$P(h|\lambda,T) = \exp(-\lambda T)\frac{(\lambda T)^h}{h!} \quad (1)$$

where $h = 0, 1, 2, \ldots$ and $\lambda > 0$, $T > 0$.

[illegible]

The Regression model

For each observation n_i, $i = 1, 2, \ldots, N$, let $Z_i = \log(\lambda_i)$, where λ_i is the true Poisson intensity. The link between the intensity and the covariates is expressed as $Z_i = X_i\beta + \varepsilon_i$, where $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_K)'$ is a vector of parameters, ε_i is assumed to be independent and identically distributed (i.i.d.) and normally distributed with zero mean and variance σ^2, and $X_i = (1, X_{i1}, X_{i2}, \ldots, X_{iK})$ is the explanatory vector for n_i.

In the vector form, this model can be formulated as below:

$$P(\mathbf{h}|\mathbf{Z})=\prod_{i=1}^{N}P(h_i|Z_i),\ \text{where}\ h_i|Z_i \sim Poisson(e^{Z_i})$$

$$\mathbf{Z}|\boldsymbol{\beta},\sigma^2,\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_N),\ \text{where, specifically}$$

$$\mathbf{X}=[X_1,X_2,\ldots,X_N]'$$

$$X_i=[1,SST_i,PW_i,VOR_i,VWS_i],\ i=1,2,\ldots,N$$

$$\boldsymbol{\beta}=[\beta_0,\beta_1,\beta_2,\beta_3,\beta_4]' \qquad (2)$$

Based on Bayes' theorem, it is obvious that

$$P(\mathbf{Z}\,|\,\mathbf{h},\boldsymbol{\beta},\sigma^2) \propto P(\mathbf{h}\,|\,\mathbf{Z},\boldsymbol{\beta},\sigma^2)P(\mathbf{Z}\,|\,\boldsymbol{\beta},\sigma^2)$$
$$= P(\mathbf{h}\,|\,\mathbf{Z})P(\mathbf{Z}\,|\,\boldsymbol{\beta},\sigma^2) \qquad (3)$$

Substituting the probability model (2) into (3), ignoring the constant part, yields the conditional posterior distribution for a Poisson model

$$P(\mathbf{Z}\,|\,\mathbf{h},\boldsymbol{\beta},\sigma^2) \propto \frac{1}{\sigma^N}\prod_{i=1}^{N}\exp\left[(-e^{Z_i}) + Z_i h_i - \frac{1}{2\sigma^2}(Z_i - \mathbf{X}_i\boldsymbol{\beta})^2\right]$$

(4)

Since we do not have any credible prior information for the coefficient vector β and the variance σ^2, it is reasonable to choose the non-informative prior. In formula, it is (Gelman et al., 2004, p. 355)

$$P(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}$$

With the new observed predictor set $\tilde{X} = [1, \tilde{X}_{i1}, \tilde{X}_{i2}, \ldots, \tilde{X}_{iK}]$, if we have the posterior distribution of the parameters, the predictive distribution for the latent variable $\tilde{Z}$ and TC counts $\tilde{h}$ will be

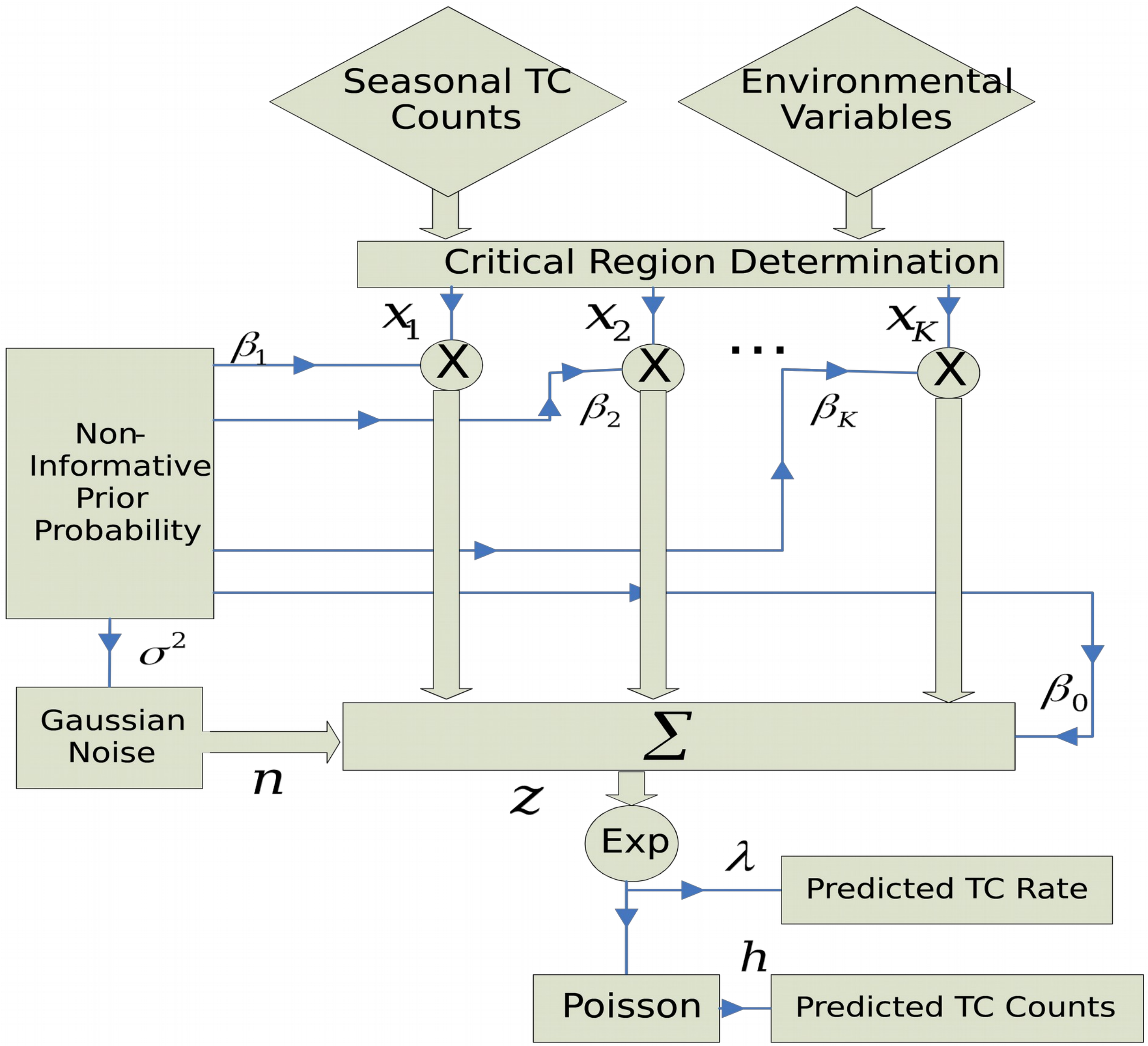
$$P(\tilde{Z} \mid \tilde{X}, \mathbf{X}, \mathbf{h}) = \int\int_{\boldsymbol{\beta}, \sigma^2} P(\tilde{Z} \mid \tilde{X}, \boldsymbol{\beta}, \sigma^2) P(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{h}) d\boldsymbol{\beta} d\sigma^2 \quad (5a)$$

$$P(\tilde{h} \mid \tilde{X}, \mathbf{X}, \mathbf{h}) = \int_{\tilde{Z}} \frac{\exp(-e^{\tilde{Z}} + \tilde{Z}\tilde{h})}{\tilde{h}!} P(\tilde{Z} \mid \tilde{X}, \mathbf{X}, \mathbf{h}) d\tilde{Z} \quad (5b)$$

We design a Gibbs sampler, which has $P(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{h})$ as its stationary distribution, and then we can use an alternative approach to integrate (5a) by

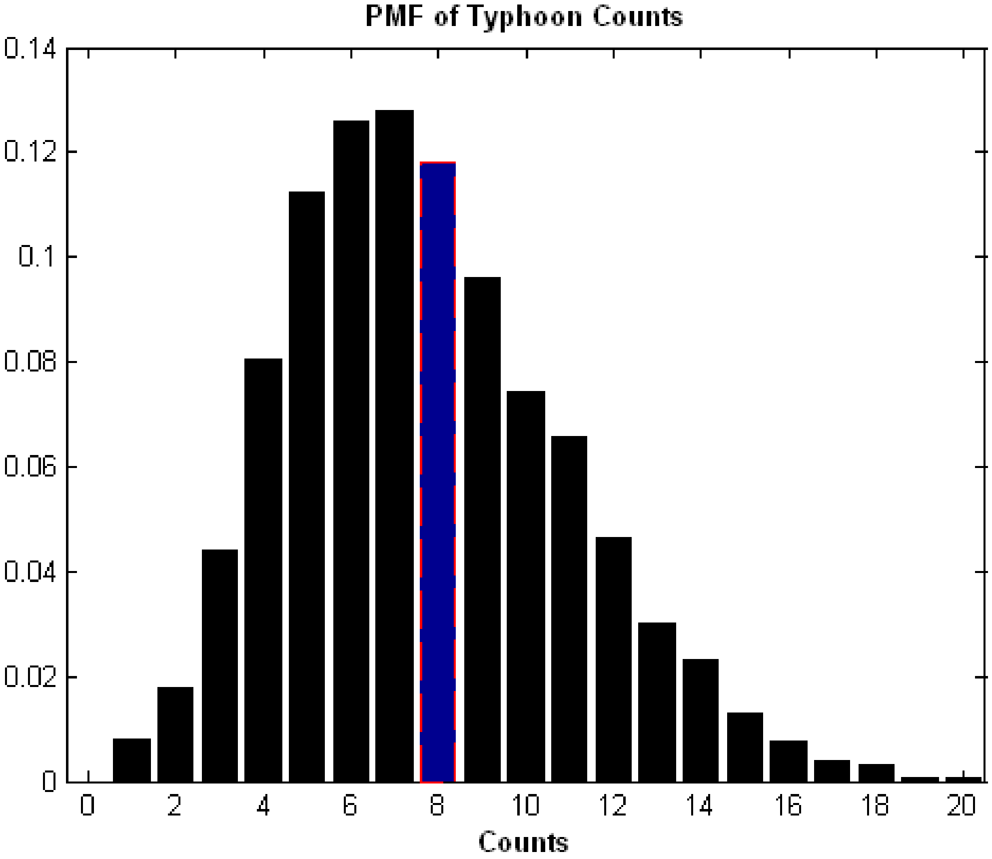
$$P(\tilde{Z} \mid \tilde{X}, \mathbf{X}, \mathbf{h}) = \frac{1}{L} \sum_{i=1}^{L} P(\tilde{Z} \mid \tilde{X}, (\boldsymbol{\beta}, \sigma^2)^{[i]}) \qquad (6)$$

where $(\boldsymbol{\beta}, \sigma^2)^{[i]}$ is the i-th sampling from the Gibbs sampler after the burn-in period.



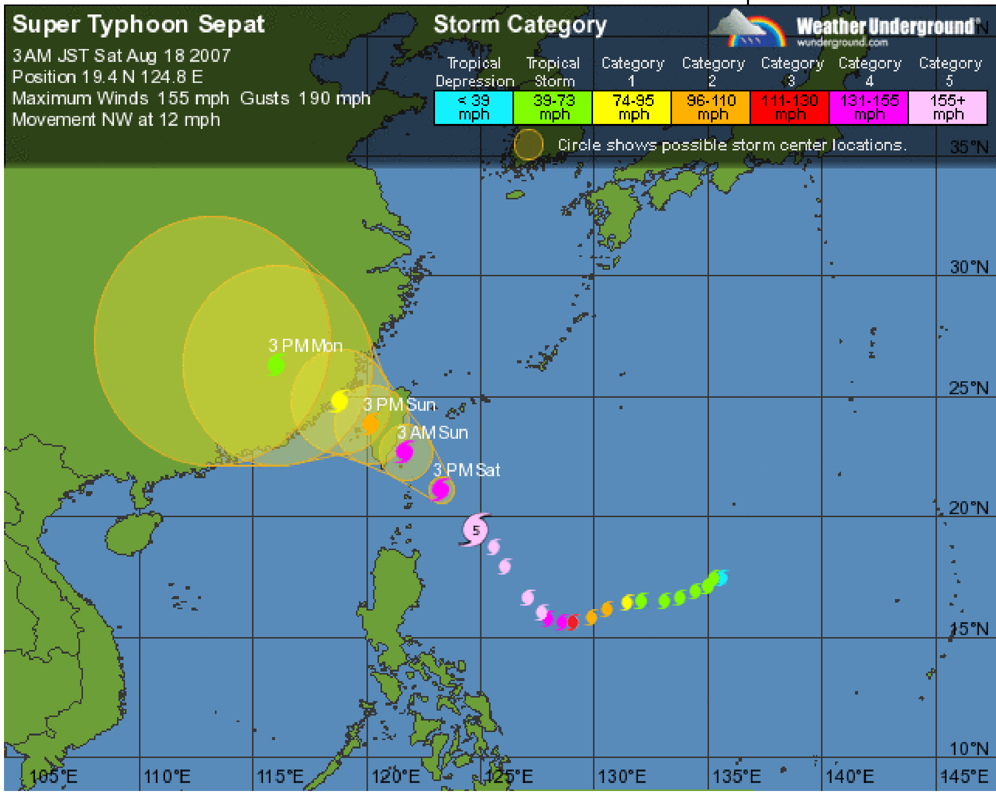
Seasonal TC Counts
Environmental Variables
Critical Region Determination
x_1
x_2
x_K
β_1
β_2
β_K
X
X
X
...
Non-Informative Prior Probability
σ^2
Gaussian Noise
n
Σ
β_0
z
Exp
λ
Predicted TC Rate
h
Poisson
Predicted TC Counts

Forecasting a busy year (2004)

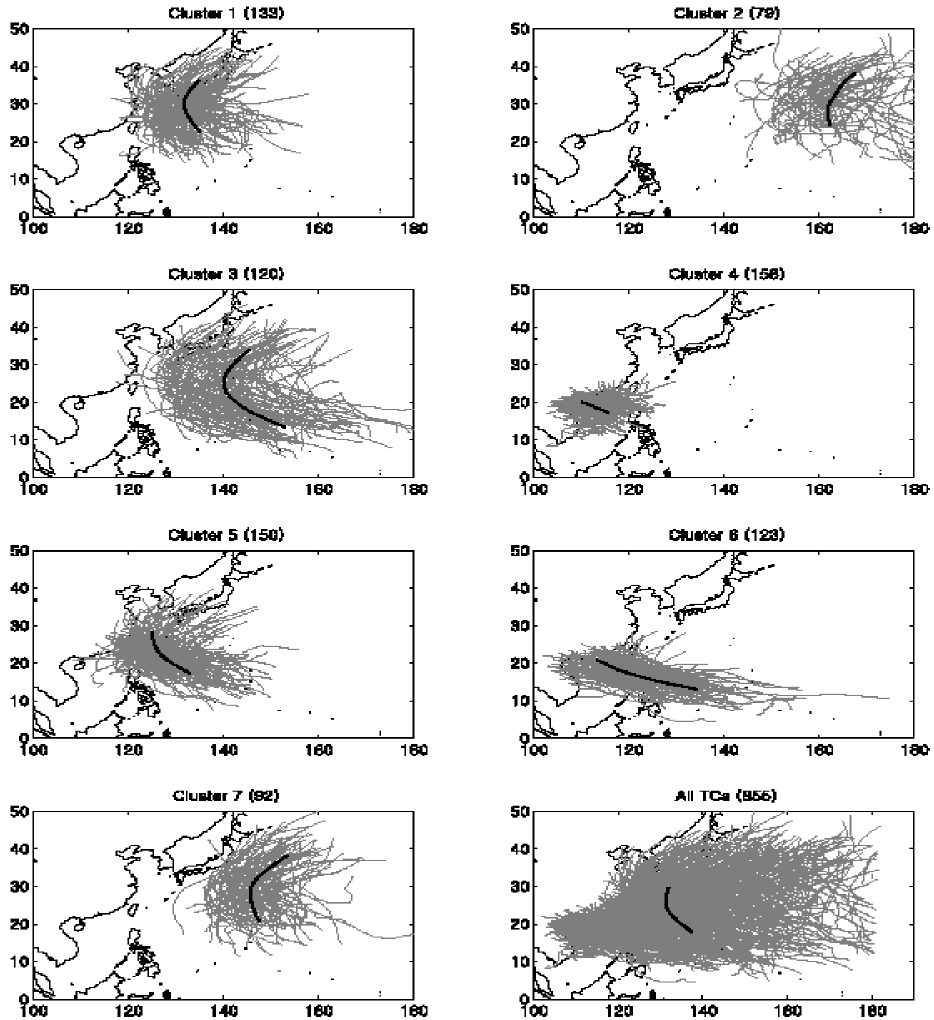


Part II: A Track-type based Approach

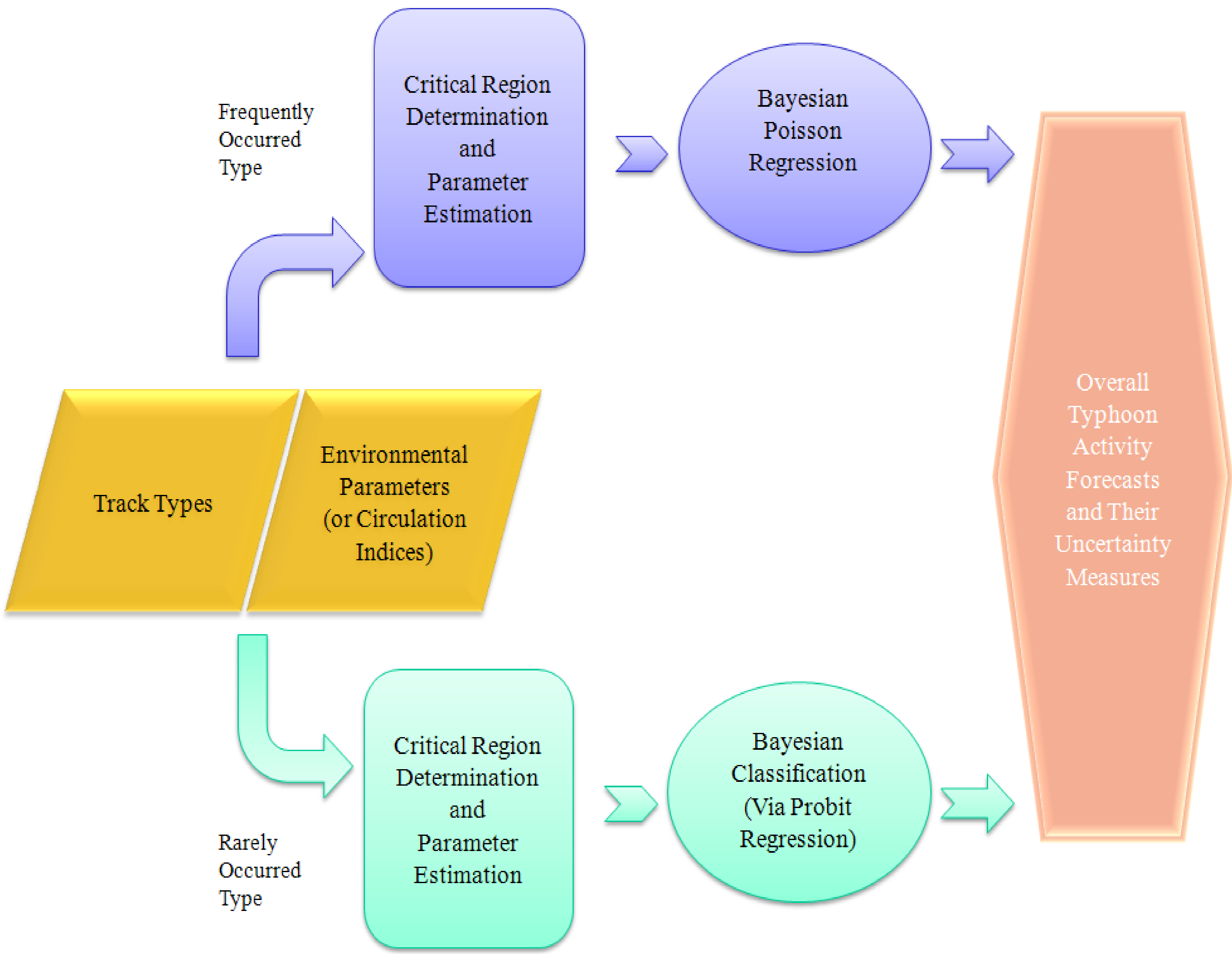
- A vector EOF analysis for TC tracks
- Fuzzy clustering of TC tracks (Harr and Elsberry, 1995)



SNU group


Cluster 1 (133)
Cluster 2 (79)
Cluster 3 (120)
Cluster 4 (158)
Cluster 5 (150)
Cluster 6 (123)
Cluster 7 (92)
All TCs (855)
0
10
20
30
40
50
100
120
140
160
180

Year	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Total
1997	0	0	1	0	1	0	0	2
1998	1	0	0	1	2	0	0	4
1999	1	0	0	0	0	1	0	2
2000	1	0	0	0	4	1	0	6
2001	1	0	0	0	4	1	0	6
2002	0	0	1	0	3	0	0	4
2003	0	0	0	0	5	1	0	6
2004	0	0	0	1	6	1	0	8
2005	0	0	0	0	5	0	0	5
2006	0	0	0	0	4	1	0	5


Frequently Occurred Type
Critical Region Determination and Parameter Estimation
Bayesian Poisson Regression
Track Types
Environmental Parameters (or Circulation Indices)
Overall Typhoon Activity Forecasts and Their Uncertainty Measures
Rarely Occurred Type
Critical Region Determination and Parameter Estimation
Bayesian Classification (Via Probit Regression)

Summary

- Two methods for predicting seasonal typhoon activity are introduced (Basin-wide to regional, deterministic to probabilistic).

  A state-of-the-art hierarchical Bayesian system (i.e., Poisson or probit regression), rooted on the track patterns, is currently being developed to provide probabilistic forecasts for typhoons near Taiwan.

  It may be possible to forecast mean genesis locations, mean path, and landfall locations for each track types.